<u>Amendment to the Claims</u>

Please amend claims 1-3, 7, 9, 10, 11, 13, 15, 18 and 20 as follows:

<u>Listing of the claims</u>

1. (Currently Amended) A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for <u>recognizing speech</u>~~rescoring N-best hypotheses of an automatic speech recognition system~~, the method steps comprising:

generating a synthetic waveform for each of N <u>textual transcriptions</u>~~text sequences~~ <u>of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent</u>~~representing the~~ N-best <u>textual transcription</u> hypotheses <u>of the original waveform</u>~~output from a speech recognition system~~;

comparing each synthetic waveform with ~~an~~<u>the</u> original waveform decoded by the speech recognition system to determine the synthetic waveform that is closest to the original waveform; and

selecting for output the <u>textual transcription</u>~~text sequence~~ corresponding to the synthetic waveform determined to be closest to the original waveform

2. (Currently Amended) The program storage device of claim 1, wherein the instructions for performing the comparing step include instructions for performing the steps of:

aligning frames of the original waveform and frames of each synthetic waveform to a corresponding one of the N ~~text sequences~~<u>textual transcriptions</u>; and

2

calculating the distance between the original waveform and each of the synthetic waveforms based on the corresponding alignments.

3. (Currently Amended) The program storage device of claim 2, wherein the instructions for performing the comparing step further include instructions for:

retrieving feature vectors corresponding to the original waveform; and

generating feature vectors for each synthetic waveform such that the feature vectors for the synthetic waveforms are similar in structure to the feature vectors of the original waveform;

wherein the alignment is performed by time-aligning the feature vectors of the original waveform and the feature vectors of each synthetic waveform with the corresponding one of the N ~~text sequences~~textual transcriptions.

4. (Original) The program storage device of claim 2, wherein the alignment is performed using Viterbi alignment process.

5. (Original) The program storage device of claim 2, wherein the alignment is performed on a phoneme level.

6. (Original) The program storage device of claim 2, wherein the instructions for calculating the distance include instructions for performing the steps of:

calculating an individual distance between each aligned frame of the original waveform and each of the N synthetic waveforms; and

summing the individual distances of the aligned frames of the original waveform and each synthetic waveform.

7. (Currently Amended) The program storage device of claim 1, wherein the instructions for performing the comparing step include instructions for performing the steps of:

(a) setting a parameter N=1;

(b) retrieving the Nth synthetic waveform and the corresponding Nth ~~text sequence~~textual transcription;

(c) time-aligning frames of the original waveform and frames of the Nth synthetic waveform to corresponding text of the Nth ~~text sequence~~textual transcription;

(d) computing an individual distance between each corresponding aligned frame of the original and Nth synthetic waveform;

(e) summing the individual distances to compute the distance between the original waveform and the Nth synthetic waveform;

(f) determining if the computed distance is than a current best distance value;

(g) setting the current best distance value equal to the computed distance and saving the Nth ~~text sequence~~textual transcription for consideration as the final output, if the computed distance is determined to be less than the current best distance threshold;

(h) incrementing the parameter N by one; and

(i) repeating steps (b) through (h) until each of the N ~~text sequence~~textual transcriptions have been considered.

8. (Original) The program storage device of claim 7, wherein the instructions for performing the step of determining the individual distance (step d) include instructions for:

computing a mean feature vector of all feature vectors comprising each aligned frame for both the original and Nth synthetic waveform, wherein the individual distance for each aligned frame is calculated by determining a distance between each mean of the corresponding aligned frames.

9. (Currently Amended) A method for recognizing speech~~rescoring N-best hypotheses of an automatic speech recognition system~~, the method comprising the steps of:

generating a synthetic waveform for each of N textual transcriptions~~text sequences~~ of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent ~~representing the~~ N-best textual transcription hypotheses of the original waveform~~output from a speech recognition system~~;

comparing each synthetic waveform with ~~an~~the original waveform decoded by the speech recognition system to determine the synthetic waveform that is closest to the original waveform; and

selecting for output the ~~text sequence~~textual transcription corresponding to the synthetic waveform determined to be closest to the original waveform.

10. (Currently Amended) The method of claim 9, wherein the comparing step includes the steps of:

aligning frames of the original waveform and frames of each synthetic waveform to a corresponding one of the N textual transcriptions~~text sequences~~; and

calculating the distance between the original waveform and each of the synthetic waveforms based on the corresponding-alignments.

11. (Currently Amended) The method of claim 10, wherein the comparing step further includes the steps of:

retrieving feature vectors corresponding to the original waveform; and

generating feature vectors for each synthetic waveform such that the feature vectors for the synthetic waveforms are-similar in structure to the feature vectors of the original waveform;

wherein the alignment is performed by time-aligning the feature vectors of the original waveform and the feature vectors of each synthetic waveform with the corresponding one of the N textual transcriptionstext sequences.

12. (Original) The method of claim 10, wherein the step of calculating the distance includes the steps of:

calculating an individual distance between each aligned frame of the original waveform and each of the N synthetic waveforms; and

summing the individual distances of the aligned frames of the original waveform and each synthetic waveform.

13. (Currently Amended) The method of claim 9, wherein the comparing step includes the steps of:

(a) setting a parameter N=1;

(b) retrieving the Nth synthetic waveform and the corresponding Nth text sequencetextual transcription;

(c) time-aligning frames of the original waveform and frames of the Nth synthetic

waveform to corresponding text of the Nth ~~text sequence~~textual transcription;

(d) computing an individual distance between each corresponding aligned frame of

the original and Nth synthetic waveform;

(e) summing the individual distances to compute the distance between the original

waveform and the Nth synthetic waveform;

(f) determining if the computed distance is less than a current best distance value;

(g) setting the current best distance value equal to the computed distance and

saving the Nth ~~text sequence~~textual transcription for consideration as the final output, if

the computed distance is determined to be less than the current best

distance threshold;

(h) incrementing the parameter N by one; and

(i) repeating steps (b) through (h) until each of the N ~~text sequences~~textual

transcriptions have been considered.

14. (Original) The method of claim 13, wherein the step of determining the

individual distance (step d) includes the steps of:

computing a mean feature vector of all feature vectors comprising each aligned

frame for both the original and Nth synthetic waveform, wherein the individual distance

for each aligned frame is calculated by determining a distance between each means of the

corresponding aligned frames.

15. (Currently Amended) An automatic speech recognition system, comprising:

a decoder for decoding an original waveform of acoustic utterances to produce N ~~text sequence~~textual transcriptions, the N ~~text sequences~~textual transcriptions representing N-best textual transcription hypotheses of the decoded original waveform;

a waveform generator for generating a synthetic waveform for each of the N ~~text sequences~~textual transcriptions; and

a comparator for comparing each synthetic waveform with the original waveform to rescore the N-best hypotheses,

wherein N is greater than 1.

16. (Original) The system of claim 15, further comprising a feature analysis processor adapted to generate a set of feature vectors for the original waveform and generate a set of feature vectors for each of the N synthetic waveforms using a similar feature analysis process.

17. (Original) The system of claim 15, further comprising a processor adapted to process one of the original waveform, the synthetic-waveforms, and both, to compensate for speaker-dependent variations.

18. (Currently Amended) The system of claim 15, wherein the comparator comprises:

means for determining the synthetic waveform that is closest in distance to the original waveform; and

means for outputting the N ~~text sequence~~textual transcription corresponding to the synthetic waveform that is determined to be closest to the original waveform.

19. (Original) The system of claim 18, wherein the means for determining the closest synthetic waveform utilizes one of a distance score, a language model score, an acoustic model score, and a combination thereof, for determining the closest distance.

20. (Currently Amended) The system of claim 18, wherein the means for determining the closest synthetic waveform comprises:

means for aligning frames of the original waveform and frames of each synthetic waveform to a corresponding one of the N ~~text sequences~~textual transcriptions; and

means for calculating the distance between the original waveform and each of the synthetic waveforms based on the corresponding alignments.

21. (Original) The system of claim 20, wherein the frames are aligned on a phoneme level.

22. (Original) The system of claim 20, wherein the means for calculating the distance comprises:

means for calculating an individual distance between each aligned frame of the original waveform and each of the N synthetic waveforms; and

means for summing the individual distances of-the aligned frames of the original waveform and each synthetic waveform to compute the distance between the original waveform and each synthetic waveforms.